Melting of genomic DNA: predictive modeling by nonlinear lattice dynamics

Nikos Theodorakopoulos[1,2]

[1] *Theoretical and Physical Chemistry Institute, National Hellenic Research Foundation,*
Vasileos Constantinou 48, 116 35 Athens, Greece
[2] *Fachbereich Physik, Universität Konstanz, 78457 Konstanz, Germany*
(Dated: August 5, 2010)

The melting behavior of long, heterogeneous DNA chains is examined within the framework of the nonlinear lattice dynamics based Peyrard-Bishop-Dauxois (PBD) model. Data for the pBR322 plasmid and the complete T7 phage have been used to obtain model fits and determine parameter dependence on salt content. Melting curves predicted for the complete fd phage and the Y1 and Y2 fragments of the ϕX174 phage without any adjustable parameters are in good agreement with experiment. The calculated probabilities for single base-pair opening are consistent with values obtained from imino proton exchange experiments.

I. INTRODUCTION

The phenomenon of thermal DNA denaturation can be interpreted in terms of mesoscopic, one-dimensional, statistical or nonlinear lattice-dynamical models. Models of the first type, based on the equilibrium statistical mechanics of the helix-coil transition, proposed and solved long ago by Poland and Scheraga (PS) [1] have been extensively studied, refined and developed by subsequent research; this substantial effort, incorporating a large body of enthalpic data, has led to algorithms for the detailed prediction of melting profiles with a rich internal structure [2, 3]; as a result, the microscopic parametrization of this family of models is well advanced [4]. Models of the second type, first proposed by Peyrard, Bishop and Dauxois (PBD) [5–7] were largely motivated by the study of dynamical properties, in particular those associated with the local breathing of DNA which characterizes the initiation of the transcription process; in the limiting case of infinite, homogeneous chains they describe DNA denaturation as an exact, one-dimensional thermodynamic phase transition [8], whose effective (observable) order is controlled by the strength of the nonlinear base-stacking interaction [9, 10] and can be followed in detail by finite-size scaling analysis [11]. Curiously, although this type of model is uniquely suited [9, 12] to describe the sharp multistep melting process taking place at long sequences, where the effects of cooperativity are most likely to be correctly modeled with a small set of coupling parameters, comparisons of PBD-based DNA melting curves with experimental data have been confined to relatively short chains where multiphase melting is either absent [13, 14] or marginal [15–17]. As a result, one of the major potential sources of experimental validation of the PBD model remains unexplored.

In this paper I will present an analysis of melting profiles based on the standard version of the PBD model [6] for a number of specific sequences with many thousands of base pairs. I will use the parameters obtained by fitting two sets of published experimental data corresponding to different ionic contents, the pBR322 plasmid and the T7 phage melting profiles, to estimate the parameters' variation with ionic content. This procedure provides model parameters for any ionic content in the usual experimental range and thus paves the way for predicting the melting profile of any long DNA sequence. I present predicted melting profiles with no adjustable parameters and compare them with experimental data in three cases: the complete fd-phage, and the Y1 and Y2 fragments of the ϕX174 phage. The computed profiles, although not perfect, reflect accurately many of the actual complexities of sequence-dependent melting. Furthermore, the set of parameters obtained by fitting the melting profiles can be used to predict the probability of single base-pair opening at ambient temperatures. The calculated values, in the order of 1 ppm, are in line with estimates obtained from imino proton exchange measurements [18]. In summary, the PBD model seems to provide an excellent mesoscopic, dynamically motivated framework for the quantitative description of genomic DNA melting; this could be of particular use in further work where the model enjoys a comparative advantage, e.g. in extending the study of local DNA flexibility [19] to longer sequences.

II. MODEL NOTATION AND NUMERICAL PROCEDURE

The configurational energy of a chain of N base pairs is

$$H_P = \sum_{j=1}^{N-1} W(y_j, y_{j+1}) + \sum_{j=1}^{N} V_j(y_j) \qquad (1)$$

where the transverse coordinate y_j represents the separation of the two bases at the jth site, the anharmonic elastic term

$$W(y_j, y_{j+1}) = \frac{k}{2} \left[1 + \rho e^{-b(y_j + y_{j+1})} \right] (y_j - y_{j+1})^2 \quad (2)$$

models the nonlinear base-stacking interaction, and the on-site Morse potential $V_j(y_j) = D_j(1 - e^{-\alpha_j y_j})^2$ de-

scribes the combined effects of hydrogen-bonding, stacking and solvent acting on the jth base pair. I will use the same parameter values $k = 0.00045$ eV$/A^2$ for the linear part of the base stacking interaction, $b = 0.2A^{-1}$ for the inverse range of the nonlinear base stacking, and $\alpha_{AT} = 4.2A^{-1}$, $\alpha_{GC} = 6.9A^{-1}$ for the inverse ranges of the Morse potentials for all sequences discussed in this paper; ρ, the relative strength of the nonlinear base stacking will be set equal to 50, which is roughly comparable to the ratio of accepted values for the bending stiffness of double-stranded and single-stranded DNA[20]. This leaves only the depths of the Morse potential D_{AT} and D_{GC} to be determined from the data (cf. below). Furthermore, I will assume that the chain is subject to free-end boundary conditions.

Key thermodynamic quantities of interest are (i) the partition function

$$Z_N = \int_{-\infty}^{\infty} dy_1 \cdots dy_N e^{-H_P/(k_B T)} \qquad (3)$$

and (ii) the probability that the rth base pair is bound,

$$q_r = \frac{1}{Z_N} \int_{-\infty}^{\infty} dy_1 \cdots \int_{-\infty}^{\infty} dy_{r-1} \int_{-\infty}^{y_c} dy_r$$
$$\int_{-\infty}^{\infty} dy_{r+1} \cdots \int_{-\infty}^{\infty} dy_N e^{-H_P/(k_B T)} \qquad (4)$$

where k_B is the Boltzmann constant, T the temperature and y_c is an appropriately chosen displacement which distinguishes the open ($y > y_c$) from the closed ($y < y_c$) state of a base pair. Unless otherwise stated, the choice will be $y_c = 2A$.

As has been extensively discussed in the literature (e.g. [12]) the partition function (3) diverges because of the flat top of the Morse potential; however, this divergence - ultimately responsible for the occurrence of an exact phase transition in the thermodynamic limit - is restricted to the disordered phase[21]; as a result, (4) is always well defined - and can be shown to be numerically stable, i.e. independent of any upper cutoff of the integrations - for sufficiently long chains [22].

Numerical work was based on expanding [12, 24] each factor in the partition function in terms of the complete set of eigenstates of the integral equation

$$\int_{-\infty}^{\infty} dy' K(y, y') \phi_\nu(y') = \Lambda_\nu \phi_\nu(y) \qquad (5)$$

with $K(y, y') = e^{-[W(y,y')+V_{AT}(y)/2+V_{AT}(y')/2]/k_B T}$ serving as a "reference" kernel. This transforms the multidimensional integrals in Eqs. (3) and (4) into matrix products in the reduced space of eigenstates of (5) - typically of dimension 30-100 if one keeps only eigenstates with $\Lambda_\nu/\Lambda_0 > 10^{-8}$; the resulting matrix products can be calculated quite rapidly if one stores the intermediate results. Numerical computations have been performed by substituting the lower and upper limits of integration by $-1.5A$ (owing to the rapid vanishing of the kernel

which originates in the repulsive core of the Morse potential) and $300A$, respectively; extending the region of integration in either direction does not produce any detectable change in the resulting melting profiles. Eq. 5 has been discretized using Gauss-Legendre quadratures with a mesh of 1200 points. Further numerical details will be reported elsewhere.

III. RESULTS

A. Melting profiles

The fraction of unbound ("melted") base pairs is given by

$$\theta = 1 - \frac{1}{N} \sum_{i=1}^{N} q_i \quad . \qquad (6)$$

Fig. 1 shows calculated and experimental differential melting curves $d\theta/dT$ vs. T for (i) the plasmid pBR322 (4361 bps, GC content 53.8 %, Na$^+$ ion concentration $c_0 = 0.075M$, $D_{AT}^0 = 0.1255$ eV, $D_{GC}^0 = 0.1655$ eV), and (ii) the the T7 phage (39937 bps, 48.4% GC content, Na$^+$ ion concentration $c = 0.0195M$, $D_{AT} = 0.1205$ eV, $D_{GC} = 0.1619$ eV). The calculated melting curves, although not in perfect agreement with the experimental ones, reflect to a considerable extent the latter's complexities. Note in particular the multipeak structure on the left panel, and the correct overall shape of the melting curve on the right panel. The superimposed melting map shown in the left panel (melting temperatures, in steps of 0.5 K, uniquely defined for each base pair via $p_i[T_m(i)] = 1/2$) exhibits the distinct vertical regions known from statistical (PS-type) theories of DNA denaturation to coincide with peaks of the melting profile and to characterize cooperative melting of large domains, extending over hundreds of base pairs.

A careful analysis of the critical behavior of pure AT and GC sequences reveals essentially linear dependence of the melting temperature on D in the range of interest. Since T_m is known [23] to depend logarithmically on c, it is natural to assume that

$$D_\sigma = D_\sigma^0 + \kappa_\sigma \log(c/c_0) \quad , \qquad (7)$$

where $\sigma = AT, GC$, and use the two sets of Morse depth parametes to extract estimates of the proportionality constants, $\kappa_{AT} = 0.00855$ eV and $\kappa_{GC} = 0.00615$ eV, respectively [25]. This allows the calculation of D_{AT} and D_{GC} at any salt concentration, and therefore, in principle, the prediction of melting behavior for any long DNA sequence.

I have tested the validity of the above procedure in three cases. Fig. 2 compares predicted to experimental melting profiles for the Y1 and Y2 fragments of the ϕX174 phage. The Na$^+$ ion concentration is $c = 0.195M$ and the corresponding Morse depths, based on
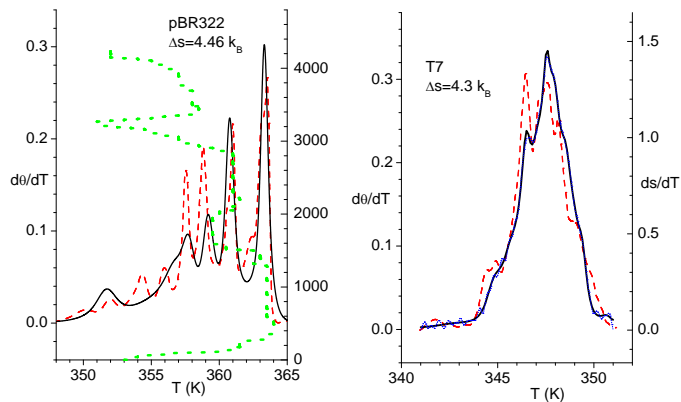
FIG. 1. (color online) *Left panel*: left y-axis, melting curve for the plasmid pBR322 (4361 bps), solid line PBD model calculation, dashed red line experimental data redrawn from [3]; superimposed (dotted line) is the melting map, to be read as melting temperature vs. site (right vertical axis); *right panel*: computed melting profile (solid line, left y-axis) for the T7 phage (39937 bps); experimental data redrawn from [29] (dashed red line); also shown (dotted curve, almost indistinguishable from the melting curve, right y-axis) is the temperature derivative of the entropy per site.

(7), $D_{AT} = 0.12905$ and $D_{GC} = 0.16805$ respectively. Fig. 3 compares predicted to experimental melting profiles for the fd-phage (6408 bps, 40.9% GC content). The Na^+ ion concentration was [26] $c = 0.0195M$, therefore the Morse depths used are the same those of the T7-phage (cf. above). In all three cases, PBD-based calculations with no adjustable parameters successfully predict melting temperatures and profiles in considerable detail.
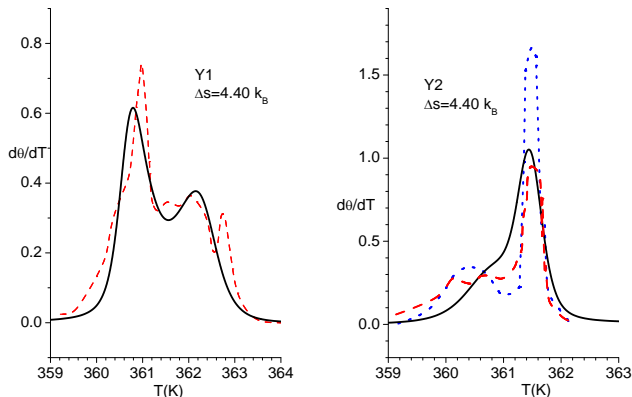

FIG. 2. (color online) *left panel*: Melting curves for Y1 fragment (2746 bps) of the ϕX174 phage, solid line PBD model prediction, dashed line experimental curve redrawn from [2]; *right panel*: same for Y2 fragment (1695 bps), experimental data redrawn from [2] based on two different experiments (dashed line [28], dotted line [27]).

B. Longer sequences

The numerical method used here can in principle deal with much longer sequences. The right panel of Fig. 3 shows the theoretical melting profile obtained for the full genome sequence of the bacterial endosymbiont *Carsonella ruddii* (159662 bps, GC content 16.6%, shortest genome of any characterized bacteria [30]) using - somewhat arbitrarily - the same parameters as in the Y1 and Y2 fragments (cf. above). The transition has a temperature width of almost 20 K; this is unusually broad for such a large system and originates in large scale fluctuations of locally averaged GC-content.

C. Melting Entropy

The entropic signature of the melting transition can be followed in some detail by looking at the temperature derivative ds/dT of the entropy per site $s = [k_B/N]d(T \ln Z_N)/dT$. The right panel of Fig. 1 shows that the proportionality relationship

$$\frac{ds}{dT} = \Delta s \frac{d\theta}{dT} \qquad (8)$$

holds to a great degree of accuracy (the two curves being practically coincident), with Δs being defined as the entropy of melting. Within the PBD model framework, calorimetry and base-pair sensitive methods (e.g. UV absorption) deliver essentially identical profile information. Numerical integration gives values $\Delta s = 4.46, 4.35, 4.41 k_B$, for the pBR322 plasmid, the fd and T7 phages respectively, and 4.40 k_B for the Y1 and Y2 fragments; although considerably lower than typical experimental values [3] of $12k_B$ (corresponding to 24 cal/mol(bp)/K), these values represent a significant improvement over previous estimates, around $1k_B$, obtained [10] using oligomer-based [13] parameters.

D. Opening of individual base pairs

Base pair opening has been carefully investigated [18] by making use of the imino proton exchange technique. The experiments suggest (i) a closed base pair lifetime in the order of a few milliseconds and (ii) equilibrium constants between open and closed states corresponding to a fraction of open base pairs of the order of 1 ppm (at room temperature). Since it is not immediately obvious what values of $y > y_c$ correspond to the open state in the imino proton experiment, I have plotted in Fig. 4 the melting fraction as a function of temperature for a range of ρ values and two choices of critical amplitude y_c, equal to $2A$ and $5A$, respectively. The results indicate that the choice $\rho = 50$, $y_c = 5A$ is consistent with the conclusions of the imino proton exchange experiments. It should of course be noted that the exact choice of y_c does not affect the
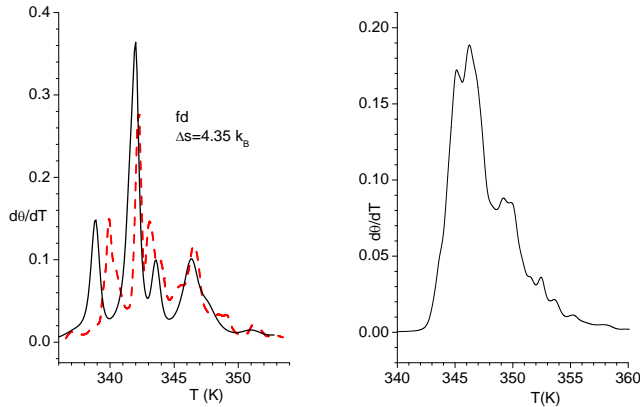
FIG. 3. (color online) *Left panel:* Melting curve for the fd-phage (6408 bps), PBD model prediction (solid line) vs experiment (dashed line, redrawn from[26]). *Right panel:* Theoretical (PBD) melting profile of Carsonella Ruddii (159662 bps).

overall shape or position of the melting profile (except for a sharpening of the edges). Nonetheless, it is reassuring to confirm that a reasonable choice of y_c, when applied to the PDB model with the above (differential melting curve - based) parametrization leads to room temperature open base-pair populations which are in agreement with an entirely different class of experiments.
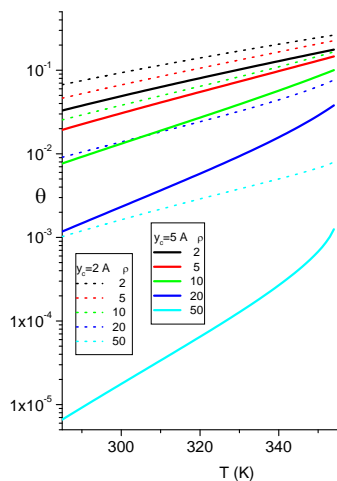

FIG. 4. (color online) Fraction of unbound base pairs as a function of temperatures calculated for the T7-phage for $\rho = 2, 5, 10, 20, 50$ and $y_c = 2A$ (dotted lines), $y_c = 5A$ (solid lines). Only the lowest curve ($\rho = 50, y_c = 5A$) leads to results consistent with estimates $\theta \sim 10^{-5} - 10^{-6}$ obtained from imino proton exchange measurements at room temperature [18].

IV. CONCLUDING REMARKS

In summary, the nonlinear lattice dynamics (PBD) approach to the statistical physics of DNA denaturation, in addition to elucidating the more abstract, qualitative aspects of the phase transition, has been shown to provide a powerful tool for the detailed description of quantitative features of melting in long genomic sequences.

[1] D. Poland and H. A. Scheraga, J. Chem. Phys. **45**, 1464 (1966); M. E. Fisher, J. Chem. Phys. **45**, 1469 (1966).
[2] R.M. Wartell and A.S. Benight, Physics Reports **126**, 67 (1985).
[3] S.G. Delcourt and R.D. Blake, J. Biol. Chem. **266**, 15160 (1991).
[4] R. Blossey and E. Carlon, Phys. Rev. E **68**, 061911 (2003).
[5] M. Peyrard and A.R. Bishop, Phys. Rev. Lett. **62**, 2755

(1989).
[6] T. Dauxois, M. Peyrard and A. R. Bishop, Phys. Rev. E **47**, R44 (1993).
[7] T. Dauxois and M. Peyrard, Phys. Rev. E **51**, 4027 (1995).
[8] T. Dauxois, N. Theodorakopoulos and M. Peyrard, J. Stat. Phys. **107**, 869 (2002).
[9] D. Cule and T. Hwa, Phys. Rev. Lett. **79**, 2375 (1997).
[10] N. Theodorakopoulos, T. Dauxois and M. Peyrard, Phys.

Rev. Lett. **85**, 6 (2000).

[11] M. Barbi, S. Lepri, M. Peyrard and N. Theodorakopoulos, Phys. Rev. E **68**, 061909 (2003).

[12] Y-L Zhang, W-M Zheng, J-X Liu, Y. Z. Chen, Phys. Rev. E **56**, 7100 (1997).

[13] A. Campa and A. Giansanti, Phys. Rev. E **58**, 3585 (1998).

[14] G. Weber, N. Haslam, N. Whiteford, A. Pruegel-Bennett, J. W. Essex and C. Neylon, Nature Phys. **2**, 55 (2006).

[15] M. Peyrard, S. Cuesta-Lopez and D. Angelov, J. Phys.: Condens. Matter **21**, 034103 (2009).

[16] A. Montrichok, G. Gruner and G. Zocchi, Europhys. Lett. **62**, 452 (2003).

[17] S. Ares, N. K. Voulgarakis, K. O. Rasmussen,and A. R. Bishop, Phys. Rev. Lett. **94**, 035504 (2005).

[18] M. Gueron, M. Kochoyan, J-L Leroy, Nature **328**, 89 (1987).

[19] G. Weber, J. W. Essex and C. Neylon, Nature Phys. **5**, 769 (2009).

[20] C.G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante, Proc. Nat. Acad. Sci. **94** 6185 (1997); S. V. Kuznetsov, Y. Shen, A. S. Benight, A. Ansari , Bioph. J. **81**, 2864 (2001).

[21] N. Theodorakopoulos, Phys. Rev. E **68**, 026109 (2003).

[22] For shorter chains q_r can be properly defined in the double-stranded ensemble, which ensures that at least one base pair is bound, cf. recent comments by (i) M. Sanrey and M. Joyeux, Phys. Rev. Lett. **102**, 029601 (2009) and (ii) T.S. van Erp, S. Cuesta-Lopez, J-H. Hagmann and M. Peyrard, Europhys. Lett. **85**, 68003 (2009).

[23] R.D. Blake and S.G. Delcourt, Nucl. Acids Res. **26**, 3323 (1998).

[24] N. Theodorakopoulos, Phys. Rev. E **77**, 031919 (2008).

[25] It may be noted here that the inequality $\kappa_{AT} > \kappa_{GC}$ is in line with the well known property [23] that the spread $T_m(GC) - T_m(AT)$ decreases with increasing salt concentration.

[26] A. Wada, H. Tachibana, O. Gotoh and M. Takanami, Nature **263**, 439 (1976).

[27] H. Tachibana, S. Ueno-Nishio, O. Gotoh and A. Wada, J. Biochem. **92**, 623 (1982).

[28] M. P. Perelroyzen, V. I. Lyamichev, A.Y. Kalambet, L.Y. Lyubchenko and A. D. Volgodskii, Nuc. Acids Res. **9**, 4043 (1981).

[29] Yu. L. Lyubchenko, M. D. Frank-Kamenetskii, A. V. Vologodskii, Yu. S. Lazurkin and G.G. Gause Jr., Biopolymers **15**, 1019 (1976).

[30] A. Nakabachi, A. Yamashita, H. Toh, H. Ishikawa, H.E. Dunbar, N.A. Moran, and M. Hattori, Science **314**, 267 (2006).